

Defragmentation-as-a-Service (DaaS): How beneficial is it?

Sandeep Kumar Singh, Wolfgang Bziuk, and Admela Jukan

Technische Universität Carolo-Wilhelmina zu Braunschweig, Germany

{sandeep.singh, w.bziuk, a.jukan}@tu-bs.de

Abstract: We analyze Defragmentation-as-a-Service (DaaS) in elastic optical networks and show that the positive effect of defragmentation depends on the rate at which it is performed, load (or, call arrival rates) and the available resources.

1. Introduction

The concept of continuity is natural when it comes to service, otherwise its counterpart, i.e. discontinuity (or fragmentation) can hurt the system performance. Fragmentation is an age old problem in storage devices, and recently it also was shown to affect the network performance in elastic optical networks (EON). In EON, spectrum can be sliced to a smaller units (e.g., 6.25 GHz) to suit the diverse applications. While serving heterogeneous application demands for bandwidth, setting up and termination of connection requests can cause spectrum fragmentation of an elastic optical link (EOL), which results in higher blocking probability. To alleviate this problem, network defragmentation, - similar to the disk defragmentation, can be used to reconfigure some of the connections such that largest possible free resource block are created for the future requests. However, reconfiguration of a connection causes disruption of its traffic for the duration of retuning time of transceivers and reconfiguration time of the optical switches. One can avoid reconfiguration time of switches by making the connection over new frequency slots (FS) before breaking it from its existing path, however migration (retuning) will still cause some interruption.

In this paper, we coin the term Defragmentation-as-a-Service (DaaS) in elastic optical networks, and analyze the positive effect of defragmentation (DF) process as a function of the rate at which it is performed, load (or, call arrival rates) and the available resources. We also analyze the impact of DF rate on the blocking probability for different spectrum allocation (SA) policies, such as first-fit (FF) and random-fit (RF) [1–3]. To understand the effect of defragmentation, we model an EOL using a multi-class continuous-time Markov chain (CTMC), and include DaaS states in addition to the normal service¹ states. We have the following assumptions for our model: i) the DF process is called when a request is blocked due to the fragmentation of a link; ii) all existing requests are interrupted during DF period, iii) all incoming requests are blocked during the DF period, and iv) system returns to normality, i.e. services of existing requests are resumed when DF is completed. Although assumptions, particularly (ii) and (iv), are stronger, traffic interruption of some (but not all) existing connections does happen during reconfiguration in the real system. Note that the DaaS states are blocking states for all types of connections, hence if DF is performed at lower rate then we would experience an increase in overall blocking probability (BP). Using our model, we show that DF reduces the overall call blocking probability only if: a) it is performed at the higher rates as compare to the normal service rate of the calls, b) load on system is moderate (i.e., call arrival rates are not high), and c) resources (bandwidth) are sufficient.

2. Proposed Model: Defragmentation-As-a-Service

We adopt a multi-class CTMC model to analyze the blocking probability of EOL. Here, we assume that a DaaS process is triggered when a request is blocked due to fragmentation of an EOL, and not due to the unavailability of resources. To model the reconfiguration of the occupied FS, we introduce DaaS states. We assume that both inter-arrival and call holding times of class- k requests are independent and exponentially distributed with average rates λ_k and μ_k , respectively. In Fig. 1(a), let us consider an example of an EOL link with $C = 6$ FS to illustrate the idea. In Fig. 1(b) and (c), we consider two classes of requests with demands 2 and 3 consecutive slots for RF and FF SA policies, respectively and show some of the transitions, due to the arrivals and departures, leading to a DaaS state. In contrast to FF, which allocates first available consecutive slots to a request, RF randomly allocates one of the possible assignments. In Fig. 1(b), under the RF method, consider a case where a new request (say R_1) of class $k = 1$ arrives in state S_1 with

¹We refer to *normal service* for the case of service which does not employ defragmentation

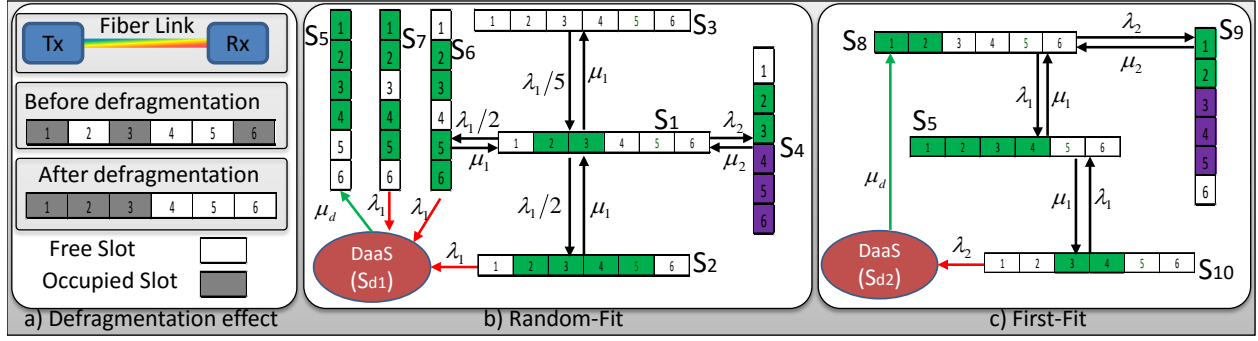


Fig. 1. DaaS: its effect, and state transitions leading to defragmentation for RF and FF SA policies.

demand $d_1 = 2$. There are $a(S_1, k) = 2$ different ways to occupy the 3 free slots of state S_1 , thus transitions ($S_1 \rightarrow S_2$ and $S_1 \rightarrow S_6$) occurs with rates $\lambda_1/2$. Furthermore, when a class-1 request (two slots) arrives, blocking of the request due to fragmentation occurs in the set of states $FB(1) = \{S_2, S_6, S_7\}$. These blocking events will trigger the transition from states $S_i \in FB(1)$ to the DaaS state S_{d1} (red arrows). During the exponentially distributed DaaS time T_{DF} with mean $1/\mu_d$, arriving calls are blocked. Furthermore, services of existing connections effected by spectrum reconfiguration are interrupted during T_{DF} . Finally, system returns to a non-fragmented state (green transition to S_5), where services of the existing requests are resumed. Due to the memoryless property of the exponential distribution, the remaining call holding time is again exponentially distributed with mean $1/\mu_k$. Thus, transitions due to call terminations are modeled with rates μ_k . The global balance equations (GBE) for the RF policy are presented in Table 1. Eq.(1a) gives the GBE

Table 1. Notation, description, and definition of key formulae of the model

K : Number of call classes.	$A(S_i, k) = \text{false}$, if a class - k request is blocked in S_i ; <i>true</i> otherwise.
λ_k (μ_k) : Arrival (service) rate of class - k calls;	$L(S_i, k) = \text{true}$, if a class - k request is present in S_i ; <i>false</i> otherwise.
$1/\mu_d$: Mean defragmentation time	$T^+(S_j, S_i, k) = \text{true}$, if transition from S_j to S_i is possible due to allocation in S_i (deallocation in S_j);
d_k : Number of slots request by class - k calls	<i>false</i> otherwise.
$\mathbf{n} \equiv (n_1, n_2, \dots, n_K)$, where n_i is the number of class i calls	\tilde{S}_i : Shifted spectrum pattern of S_i as a result of DF, $i = 1, \dots, N_{SA}$
$n_k(S_i)$: # of class - k calls in S_i	$\delta(S_i, S_{d_v})$: Comparison function for normal and DaaS states
$\mathbf{n}(S_i) \equiv (n_1(S_i), \dots, n_K(S_i))$: Realisation of \mathbf{n} in state S_i	$\delta(S_i, S_{d_v}) = 1$ if \tilde{S}_i and S_{d_v} are identical, 0 otherwise
$a(S_i, k)$: # of different ways class - k call can be allocated in S_i .	$RB(k)$: Set of blocking states for class - k due to lack of free resources,
FS_l : Size of the l^{th} fragment of free slots	$RB(k) = \{S_i d_k > C - \sum_j d_j n_j(S_i), k = 1, \dots, K, i = 1, \dots, N_{SA}\}$
$fr(S_i, k)$: Fragmentation indication function	DF : Set of states used for defragmentation
$= \begin{cases} 1 & \text{if } d_k > \max_l FS_l \text{ and } d_k \leq C - \sum_j d_j n_j(S_i) \\ 0 & \text{otherwise} \end{cases}$	$DF = \{S_{d_v} S_{d_v} \equiv \tilde{S}_i, \forall S_i \in \bigcup_k FB(k), v = 1, \dots, N_D\}$
$FB(k)$: Set of blocking states for class - k due to fragmentation	
$FB(k) = \{S_i fr(S_i, k) = 1, i = 1, \dots, N_{SA}\}, k = 1, \dots, K$	

$$\left(\sum_{k=1, A(S_i, k)} \lambda_k + \sum_{k=1, L(S_i, k)} n_k(S_i) \mu_k + \sum_{k=1} fr(S_i, k) \lambda_k \right) \pi(S_i) = \sum_{j=1, j \neq i}^{N_{SA}} \left(\sum_{k=1, T^+(S_j, S_i, k)} \frac{\lambda_k}{a(S_j, k)} + \sum_{k=1, T^-(S_j, S_i, k)} \mu_k \right) \pi(S_j) + \mu_d \sum_{S_{d_v} \in DF} \delta(S_i, S_{d_v}) \pi(S_{d_v}), \quad i = 1, \dots, N_{SA} \quad (1a)$$

$$\mu_d \pi(S_{d_v}) = \sum_{k=1} \lambda_k \sum_{S_i \in FB(k)} \delta(S_i, S_{d_v}) \pi(S_i), \quad v = 1, \dots, N_D \quad (1b)$$

$$P_B(k) = \sum_{S_i \in FB(k)} \pi(S_i) + \sum_{S_i \in RB(k)} \pi(S_i) + \sum_{S_{d_v} \in DF} \pi(S_{d_v}) \quad (2a)$$

$$P_B = \frac{\sum_k \lambda_k P_B(k)}{\sum_k \lambda_k} \quad (2b)$$

for normal occupancy states $S_i, i = 1, \dots, N_{SA}$, whereas (1b) gives the GBE for DaaS states $S_{d_v}, v = 1, \dots, N_D$. The stationary state probabilities $\pi = [\pi(S_1), \dots, \pi(S_{N_{SA}}), \pi(S_{d_1}), \dots, \pi(S_{d_{N_D}})]$ can be calculate by solving $\pi Q = 0$ subject to $\sum_i \pi(S_i) + \sum_v \pi(S_{d_v}) = 1$, where Q is the transition rate matrix. For example, the GBE for state S_1 is given by $(\lambda_1 + \lambda_2 + \mu_1) \pi(S_1) = \lambda_1/5 \pi(S_3) + \mu_1 (\pi(S_2) + \pi(S_6)) + \mu_2 \pi(S_4)$. The left (right) side characterizes the probability flow out of (into) S_1 . The input rate $\lambda_1/a(S_3, 1) = \lambda_1/5$ takes into account that 6 free slots of S_3 results into 5 different spectrum pattern, one of which is S_1 . For the output flow we have $n_1(S_1) = 1$, since there is only one existing class-1 call. The DaaS-state S_{d1} is reached from $S_i \in FB(1) = \{S_2, S_6, S_7\}$, which have different spectrum pattern, but results in same pattern after shifting, i.e. $\tilde{S}_i \equiv S_{d1}, i = 2, 6, 7$. Thus, the function $\delta(S_i, S_{d1}) = 1, i = 2, 6, 7$ assigns to the right DaaS-state, and $\delta(S_5, S_{d1}) = 1$ links to the final de-fragmented state. Using (1b), the GBE of state S_{d1} is

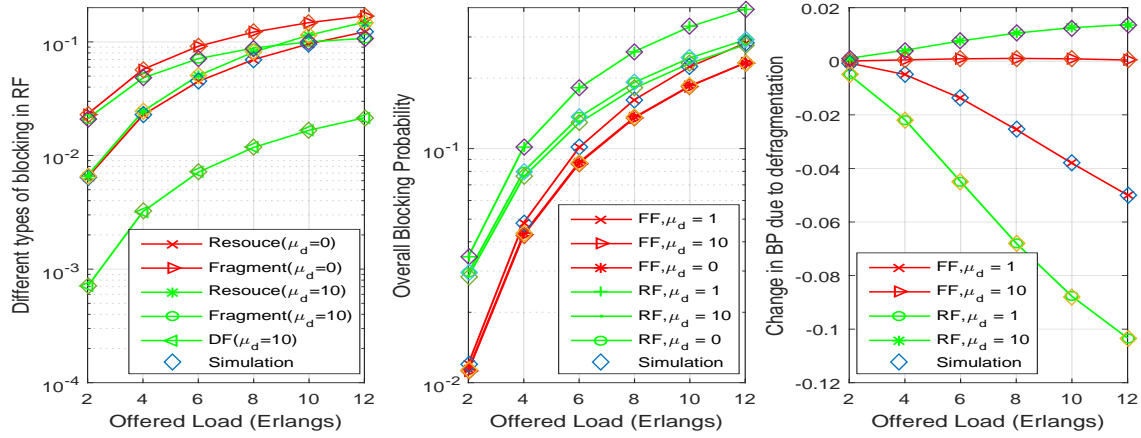


Fig. 2. Effect of DaaS when applied to RF and FF SA policies.

given by $\mu_d \pi(Sd_1) = \lambda_1 (\pi(S_2) + \pi(S_6) + \pi(S_7))$. The FF policy is also modeled by Eq.(1), taking into account that some transitions are not allowed due to the assignment strategy. Particularly, we have $a(S_i, k) = 1$. A more detailed presentation of the model is omitted for sake of brevity. The blocking probability of a class- k request is given in (2a), and the overall blocking in (2b). In our model, blocking occurs due to the fragmentation (1st term in (2a)), resource unavailability (2nd term), and also in DaaS states (last term).

3. Numerical Results

We show the analytical and verifying Monte-Carlo simulation results for an EOL having 20 FS, and three classes of requests with demand $d_1 = 4, d_2 = 6$ and $d_3 = 8$ slots. Arrival rates are uniformly distributed and mean holding time is one unit for all classes of requests. Load on the link is calculated as $\sum_{k=1}^K \frac{\lambda_k d_k}{\mu_k}$. DaaS rates of $\mu_d = 1$ and $\mu_d = 10$ are used. A regular system without DaaS is shown by $\mu_d = 0$. In Fig. 2(left), we compare the blocking parts under RF policy (Eq.(2a)) due to the effects of lack of resources, fragmentation and defragmentation. In contrast to other work, our analytical model allows us to show accurate results for the important practical range of blocking probabilities below 10%. As expected, without DaaS and for a suitable traffic load, the blocking is dominated by fragmentation and not by resource unavailability. For $\mu_d = 10$, the DaaS reduces the fragmentation blocking at cost of the resource blocking, which becomes dominant for higher load (> 6 Erl.). The additional blocking due to the DaaS is insignificantly lower (about two magnitudes lower). As expected (in Fig. 2(center)), when average defragmentation service time ($t_{DF} = 1/\mu_d$) is same as the average service time of the class- k connection (i.e., $t_{Sk} = 1/\mu_k$), then blocking probabilities are higher as compared to the system without DaaS in both FF and RF methods. To get the advantage of DaaS, t_{DF} must be lower than the average service time t_{Sk} . As shown, all parts of the blocking probability reduces significantly for RF and marginally for FF methods when we increase the DaaS rate from 1 to 10, as corresponding service time decreases proportionally. We show the change in blocking probabilities in Fig. 2(right) as the difference in blocking probabilities of with and without DaaS. First we notice, that the FF policy has overall no significant advantage of defragmentation, also for small t_{DF} . For RF, the achieved gain (change in BP) is positive for a smaller t_{DF} ($= 1/10$) for the load range shown here, however, the gain decreases at higher loads due to the negative effect of DaaS. The reason is that when arrival rates are higher, DaaS states would be called very frequently, which increases blocking.

4. Conclusion

We model for the first time the defragmentation-as-a-service and showed the correlation among DF service rate, load and the available resources. The results showed that over exploitation of DF can hurt the system performance.

References

1. Y. Yu *et al.*, "Exact performance analytical model for spectrum allocation in flexible grid optical networks," *Journal of Optical Fiber Technology*, vol. 20, no. 2, pp. 75–83, 2014.
2. H. Beyranvand *et al.*, "An Analytical Framework for the Performance Evaluation of Node-and Network-Wise Operation Scenarios in Elastic Optical Networks," *IEEE Trans. on Netw.*, vol. 62, no. 5, pp. 1621–1633, 2014.
3. Rosa, A. N. F., et al. "Statistical analysis of blocking probability and fragmentation based on Markov modeling of elastic spectrum allocation on fiber link." *Optics Communications* 354 (2015): 362-373.